Das Gute in der Informatik

Christen, Markus

Das Gute in der Informatik

Markus Christen*

1. Das Gute in der Informatik?

Die Frage nach dem Platz des Guten in der Informatik – also einer normativen Leitvorstellung, welche diese Disziplin durchdringen und prägen soll – wirkt zunächst einmal seltsam. Schliesslich assoziieren wir diese Disziplin mit der Theorie und Technologie der Informationsverarbeitung – einer nüchternen Wissenschaft mit starkem Bezug zu technischen Fragen, für deren Lösung es Computerfachleute und Software-Ingenieure braucht. Natürlich gibt es auch in dieser Disziplin methodische Standards, die «gute» von «schlechter» Informatik abgrenzen; und selbstverständlich können Informatiklösungen allerlei unethischen Zielen dienen. Doch darum soll es hier ja nicht gehen. Gefragt ist vielmehr eine grundsätzliche Verhältnisbestimmung zwischen der Informatik und dem Guten. Und versteht man Informatik als die Theorie und Praxis der Informationsverarbeitung, so wäre eine naheliegende Antwort: Die Verarbeitung von Information ist ethisch neutral; der semantische Gehalt der Information – und das Gute würde sich darin verbergen – spielt für die wissenschaftliche Praxis innerhalb dieser Disziplin erstmals keine Rolle. Das «Sein» der Information und ihrer Verarbeitung ist säuberlich vom «Sollen», das die Information ausdrücken kann, getrennt. So gesehen hätte es keinen relevanten Platz für das Gute in der Informatik und die Frage wäre beantwortet.

Hier soll aufgezeigt werden, dass diese simple Überlegung aus mehreren Gründen falsch ist. Sie ist – vielleicht wenig überraschend – praktisch falsch, vergegenwärtigt man sich schon nur die enormen Auswirkungen der Digitalisierung unserer Lebenswelt, wofür die Informatik die Leitwissenschaft darstellt. Informatik beinhaltet nicht nur das Schreiben von Programmen oder den Bau von Computern. Es geht auch um das Design ganzer Informationsflüsse in vielerlei Prozessen, was stark in den sozialen Bereich hineinfliesst und diesen formt, wie etwa die Wirtschaftsinformatik zeigt. Immer mehr soziale Sphären werden von Systemen bevölkert, die mit Sensoren, Effektoren und dazwischengeschalteter digitaler Informationsverarbeitung ausgestattet sind und die als technische Mediatoren von Produktionsprozessen und menschlichen Beziehungen wirken. Ob Wissensverwaltung, Logistiknetzwerke oder gar die Anbahnung menschlicher Partnerschaften: Digitale Technologie durchdringt unsere Lebenspraxis auf eine Weise, die es zunehmend schwierig macht, eine klare Trennlinie zwischen Anwendungen von Informatik und der Informatik selbst zu ziehen. Entsprechend ist es schwer vorstellbar, dass sich die Informatik der Frage nach dem Guten auf die oben genannte einfache Weise entziehen kann.

Die zu Beginn gegebene Antwort ist aber auch – und das ist vielleicht der interessantere Teil der folgenden Ausführungen – theoretisch falsch. Die Informatik transportiert eine ganz bestimmte Sicht auf die Welt: Sie wird verstanden als ein Konglomerat von Informationen und Informationsflüssen – und die Beherrschung der Welt drückt sich darin aus, diese Informationsflüsse messen und beeinflussen zu können. So gesehen repräsentiert die Informatik wie kaum eine andere Wissenschaft das Ideal der Messbarkeit der Welt – und damit vielleicht sogar der Messbarkeit des Guten. Entsprechend existieren theoretische Überlegungen zur Ethik der Information wie auch praktische Versuche zur Frage, wie man beispielsweise einen *moral agent* in einem Informationssystem[1] erschaffen könnte.

In den nachfolgenden Abschnitten sollen deshalb drei Skizzen zum Verhältnis zwischen der Informatik und dem Guten erstellt werden:

[1] Mit dem Begriff «Informationssystem» (*information system*) werden Systeme von Hard- und Software bezeichnet, mit deren Hilfe Informationen gesammelt, gefiltert, prozessiert, erzeugt oder verteilt werden.

* Universität Zürich, UFSP Ethik, Zollikerstrasse 117, 8008 Zürich.

E-mail: *christen@ethik.uzh.ch*
http://www.ethik.uzh.ch/en/ufsp/ma/christen.html

Markus Christen, Dr. sc. ETH, ist Leiter des «Netzwerk Ethik von Monitoring und Überwachung» am Universitären Forschungsschwerpunkt Ethik der Universität Zürich.
Studium der Philosophie, Physik, Mathematik und Biologie an der Universität Bern. Lizenziat 1996. Promotion 2006 in Neuroinformatik an der ETH Zürich. Habilitation 2015 und PD in biomedizinischer Ethik an der Universität Zürich. Visiting Scholar an der University of Notre Dame, Indiana/USA (2011 – 13).
Forschungsschwerpunkte: Empirische Ethik (Entwicklung von Messinstrumenten und von Serious Moral Games für Messung und Förderung moralischer Kompetenzen); Neuroethik (ethische Fragen von Hirninterventionen); Ethik und Technologie (Ethik der Anwendung von Informationstechnologie und Robotik in Big Data, Sicherheit, Militär und humanitären Fragestellungen); Entwicklung von Methoden in Datenanalyse und Visualisierung.
Mitglied des Ethics Advisory Board des Human Brain Project.

– Abschnitt 2 präsentiert den wohl radikalsten Gedanken: die Idee, dass Information selbst immer etwas im ethischen Sinne Gutes darstellt und dass die (bewusste) Vernichtung von Information ein zu begründender Akt darstellt – und dass demnach auch jede Form der Informationsverarbeitung im Prinzip ethisch problematisiert werden kann. Ein prominenter Vertreter dieses Gedankens ist der Informations-Philosoph *Luciano Floridi*, auf den sich diese kurze Darstellung weitgehend stützen wird.

– In Abschnitt 3 skizzieren wir ein Anwendungsfeld, das man als die Bewusstmachung der Frage nach dem Guten in der Informatik bezeichnen kann. Es geht darum zu verstehen, dass das Design von Informationstechnologien auf subtile, aber dennoch konkrete Weise die Werthaftigkeit der Sachverhalte beeinflusst, in denen diese Technologien jeweils zum Einsatz kommen. Unter dem Stichwort des *value-sensitive design* untersuchen Forscherinnen wie *Batya Friedman* und andere diese Zusammenhänge und tragen dazu bei, die Idee des Guten in der Informatik auszudifferenzieren. Bemerkenswert ist dabei, dass diese Bewusstmachung nicht nur darin bestehen kann, gewünschte Werte auf geeignete Weise in Informationssysteme «einzubauen» – etwa um mittels *nudging* die Menschen auf den Pfad der Tugend zu bringen. Diese Systeme könnten auch dazu dienen, den Nutzer selbst über seine Vorstellungen des Guten aufzuklären.

– Im Abschnitt 4 schliesslich präsentieren wir eine Forschungsrichtung, die Informationssysteme schaffen will, die selbst *moral agents* sein sollen. Angestossen wurde diese Entwicklung durch die zunehmende Automatisierung in Bereichen, wo ethische Dilemmas zu erwarten sind: in der Forschung zu autonomen Fahrzeugen oder Kampfrobotern beispielsweise. Entsprechend müssten dann diese autonom handelnden Systeme eine Vorstellung des Guten in sich tragen. Hier gelangen wir in einen Bereich, der uns teilweise weit in die Science Fiction oder in technologische Utopien führt, in denen die Informatik letztlich Systeme schafft, die uns weit überlegen und die vielleicht gar zu einer Erkenntnis des Guten befähigt sind, die uns verschlossen ist. Philosophen wie *Nick Bostrom* beschäftigen sich mit den moralischen Fragen, die eine solche «Superintelligenz» dereinst stellen könnte.

Diese drei Skizzen sollen auf exemplarische Weise die Idee des Guten in der Informatik reflektieren, ohne auf die zahlreichen Folgefragen einzugehen, die sich hier offensichtlich stellen. Dafür wird die Leserschaft auf ausgewählte Literatur verwiesen.

Bevor wir mit diesen Ausführungen beginnen, sollten wir etwas genauer umreissen, was denn mit dem Begriff des «Guten» in der Informatik überhaupt gemeint sein kann. Nachfolgend werden darunter Werte verstanden, welche die Mittel und Ziele der Informatik prägen sollen. Unter «Werten» sollen positiv besetzte Leitvorstellungen verstanden werden, für die es hinreichende Gründe gibt anzunehmen, dass sie als allgemein berücksichtigungswürdig gelten. Diese Werte haben eine längere und über die Informatik selbst hinausreichende kulturelle Geschichte und sind gewiss nicht nur an die Informatik selbst gebunden, auch wenn gewisse Werte in der Praxis der Informatik eine grössere Rolle spielen könnten als anderswo – ein Paradebeispiel ist Privatheit (*privacy*), die im Kontext des Datenschutzes eine wichtige Leitvorstellung ist. Damit ist nicht ausgeschlossen, dass Leute bezüglich dem genauen Verständnis solcher Werte unterschiedlicher Meinungen sein können. Je nach konkretem Problem, das die Informatik lösen will, dürften zudem Werte unterschiedlich gewichtet sein oder gar miteinander in Konflikt geraten. Methodische Verfahren, wie solche möglichen Wertkonflikte erkannt und wie damit umgegangen werden soll, bilden dabei ebenfalls einen Teil dessen, was man als das «Gute» in der Informatik bezeichnen kann.

2. Information als ethischer Wert[2]

Seit vielen Jahren widmen sich die *computer ethics* jenen ethischen Fragen, welche die Anwendungen von Informationstechnologien stellen. Sie stossen dabei regelmässig auf ein eigentümliches Charakteristika vieler dieser Probleme: dem Versagen moralischer Intuitionen bei jenen, welche im Cyberspace falsche Handlungen vollziehen – also sich beispielsweise illegal in Computersysteme hacken, Raubkopien und Schadsoftware erstellen oder mittels *cyberbullying* andere mobben. Der virtuelle, anonyme und distanzierende Charakter der Nutzung von Informationssystemen hat offenbar Auswirkungen auf die Einschätzung der Falschheit von Handeln mittels solcher Technologien. Für Hacker oder Cyberkriminelle haben diese Handlungen oftmals einen spielerischen Charakter – überspitzt exemplifiziert durch das «Töten» von Monstern in Computerspielen, was kaum jemand als unmoralischen Akt ansehen würde.

Der italienische Informations-Philosoph Luciano Floridi nimmt diese Beobachtung als Ausgangspunkt für seine Überlegungen: Was lässt eigentlich genann-

[2] Dieser Abschnitt fusst weitgehend auf Floridi (1999); für eine umfassendere Einführung in das Werk von Floridi wird auf Floridi (2003) verwiesen. Kritische Betrachtungen zur Informationsethik von Floridi finden sich unter anderem in einer Sondernummer der Zeitschrift *Knowledge, Technology and Politics*, Ausgabe 23(1–2) 2010: Luciano Floridi's Philosophy of Technology: Critical Reflections.

te Handlungen im Cyberspace als schlecht erscheinen? Braucht es immer einen Rückbezug auf den menschlichen Akteur, der beispielsweise als Täter seine Tugenden untergräbt oder als Opfer letztlich Schaden in der realen Welt erleidet? Was ist, wenn menschliche Akteure überhaupt nicht in das Problem involviert sind – etwa bei Massnahmen und Gegenmassnahmen im Bereich Cyberdefense, wo vieles nur schon aufgrund der raschen Reaktionszeiten automatisiert ist? Er kommt zum Schluss, dass die klassischen ethischen Theorien aus unterschiedlichen Gründen (die hier nicht weiter ausgeführt werden) nicht befriedigend erklären können, worin das Schlechte in dieser Handlungen liegt bzw. welche Idee des Guten es hier braucht, um moralische Urteile fällen zu können. Mit Blick auf neuere Entwicklungen innerhalb der Ethik selbst stellt er zudem fest, dass sich zu den klassischen anthropozentrischen Theorien zunehmend auch solche gesellt haben, die ein weiteres Bezugsfeld des Guten haben – also beispielsweise Landschaften oder das Leben an sich als etwas intrinsisch Wertvolles betrachten.

Darauf aufbauend schlägt Floridi vor, Information an sich als etwas Werthaftes zu betrachten: «Without information there is no moral action, but information now moves from being a necessary prerequisite for any morally responsible action to being its primary object» (Floridi 1999, S. 43). Damit gehen mehrere Voraussetzungen einher: Erstens werden alle Prozesse, Veränderungen, Ereignisse oder Handlungen in der Welt als Informationsprozesse verstanden. Zweitens sind die Entitäten dieser Welt «konsistente Informationspakete» (*consistent packets of information*); und die Gesamtheit dieser Entitäten, ihrer Beziehungen und der damit verbundenen Prozesse bildet die «Infosphäre». Der Gegenbegriff von Information ist Entropie, verstanden als ein semantisches Konzept (also nicht nur im Sinn der Informationstheorie als Mass für Unordnung). Entropie drückt sich aus durch die Abwesenheit von Form, Gestalt, Muster, Ausdifferenzierung oder Inhalt in der Infosphäre. Agenten sind Entitäten, welche Phänomene erzeugen, die die Infosphäre verändern – also etwas ausdifferenzieren, aber auch Entropie erzeugen, indem beispielsweise etwas zerstört wird.

Der Kern des Guten der Informationsethik ist nun dass *jede* Entität eine gewisse Würde besitzt qua ihrer Existenz. Dies schliesst digitale Repräsentationen von Information mit ein – also etwa gespeicherte Fotografien auf einer Harddisc oder Programm-Codes. Nicht etwa Leben oder Schmerz bilden die moralisch relevanten Qualitäten, sondern das Sein der Information an sich. Die sich daraus ergebenden moralischen Gesetze sind gemäss Floridi:

0. Du sollst keine Entropie in der Infosphäre erzeugen
1. Du sollst verhindern, dass Entropie in der Infosphäre erzeugt wird
2. Du sollst Entropie aus der Infosphäre entfernen
3. Du sollst Informations-Wohlfahrt (*information welfare*) fördern, indem die Infosphäre erweitert (Informations-Quantität), verbessert (Informations-Qualität) und ausdifferenziert (Informations-Vielfalt) wird.

Um dieses doch sehr abstrakt wirkende Konstrukt anzureichern, hat Floridi ein ausgefeiltes Kategoriensystem für die Infosphäre entwickelt, das die Eigenschaften der Infosphäre als Gegensatzpaare ausdrückt. So genannte modale Eigenschaften beispielsweise sind Konsistenz (die logische Möglichkeit zu existieren, Entropie drückt sich dann aus durch Inkonsistenz), Implementiertheit (d.h. die Entität könnte physisch existieren, Entropie bedeutet dann Unmöglichkeit) und Vorhandensein (d.h. die Entität existiert tatsächlich, Entropie bedeutet dann Abwesenheit). Für drei weitere Eigenschaftsklassen existieren insgesamt 27 solche Eigenschaften (ausgeführt in Floridi 1999, Tabelle 1).

Das ist gewiss eine sehr grobe Zusammenfassung eines philosophischen Systems, an dem Floridi schon viele Jahre arbeitet (mehr dazu siehe *http://www.philosophyofinformation.net/*). Mit Blick auf die Frage nach dem Guten in der Informatik sind aber folgende Anmerkungen bedenkenswert: Erstens handelt es sich um ein System, das potenziell alles in der Welt zum Gegenstand moralischer Erwägungen macht, gleichzeitig aber auch eine Intuition der Abstufung des moralischen Werts transportiert und nicht ein bestimmtes Prinzip verabsolutiert (wie beispielsweise Schutz des Leben im Biozentrismus oder Maximierung des Glücks im Utilitarismus). Es ist gewissermassen der Gehalt der Information einer Entität, die deren Schutzwürdigkeit bestimmt: Eine Skulptur hat mehr Forminformation als ein schlichter Stein, ein Säugetier ist informationstheoretisch komplexer als ein Bakterium. Zweitens hat dieses System keine intrinsisch konservative Note wie beispielsweise die Umweltethik: Schaffen von Komplexität auch mittels technischer Artefakte ist prinzipiell zuerst einmal gut, sofern damit nicht andernorts Entropie (oder das Risiko für Entropie, etwa durch Waffen) geschaffen wird. Drittens ist diese Informationsethik besonders geeignet, die spezifischen Probleme der Digitalisierung zu verstehen, beispielsweise der Schutz privater Information: Da die private Information eines Menschen diesen nicht nur beschreibt, sondern eben auch Teil des Menschen ist und damit eine eigene Würde besitzt.

Die Theorie und Praxis der Verarbeitung von Information – was ja das Kerngeschäft von Informatik ist – wird damit von einer interessanten Warte aus problematisiert: Gestellt wird die Frage, inwiefern damit Entropie geschaffen wird – also Information beispielsweise zerstört, unzuverlässig, unverfügbar oder redundant gemacht wird.

Natürlich wirft dieser Ansatz auch sehr viele Fragen auf. So ist er offensichtlich mit einem gravierenden Messproblem verbunden. Wenn der Informationsgehalt einer Entität gewissermassen deren Schutzwürdigkeit bestimmt: wie soll dieser ermittelt werden? Gewiss haben auch andere ethischen Systeme vergleichbare Probleme (etwa die Messbarkeit des Glücks im Fall des Utilitarismus), dennoch erscheint es schwer vorstellbar, den Informationsgehalt unabhängig vom (menschlichen) Beobachter zu definieren, was den Anthropozentrismus, den Floridi gerne ausschalten würde, durch die Hintertüre wieder hineinbringt. Auch ist fraglich, ob unsere moralischen Intuitionen mit diesem Ansatz mithalten können oder ob daraus nicht vielmehr eine enorme moralische Überforderung resultiert. Doch eins scheint klar: Hier wird eine radikale Vorstellung des Guten in der Informatik entwickelt.

3. Informatik-Systeme als Vermittler des Guten[3]

Weitaus näher bei den praktischen Problemen, die die Nutzung der Informationstechnologie stellt, ist der zweite hier skizzierte Ansatz: die Idee des *value-sensitive design* im Bereich von Informationssystemen, propagiert seit den frühen 1990er-Jahren von Forscherinnen wie Batya Friedman und anderen (ausführliche Informationen hierzu finden sich auf der Website *http://www.vsdesign.org/*). Ausgangslage ist die Beobachtung, dass technische Instrumente und Werkzeuge Werte vermitteln, exemplifizieren oder es auch verunmöglichen, diesen Werten Folge zu leisten. Dies gilt im besonderen Masse für informationstechnische Systeme, welche meist in komplexen, von Menschen geschaffenen Zusammenhängen zum Einsatz kommen: als Kommunikationshilfen (Handy, Skype etc.), als Planungsinstrumente (vom simplen Terminfinder Doodle bis zu komplexen Tools für die Konstruktion von Flugzeugen), für die Ideenentwicklung (kollektives Schreiben an Dokumenten etc.), für die Steuerung von technischen Systemen und so weiter. In vielen solchen Systemen finden sich *default*-Optionen, die schwer oder teilweise gar nicht veränderbar sind und die die Wahrnehmung von Werten beeinflussen. Friedman nennt das Beispiel von eingebauten Mikrofonen in Computern, die sich gar nicht mehr abschalten lassen; etwa wenn man im Fall einer Konferenzschaltung unterbrochen wird und ein privates Gespräch führen will – die Autonomie und Privatheit des Nutzers wird damit verunmöglicht. Angesichts der heutigen Überwachungsmöglichkeiten durch die Nutzung von Informationssystemen wirkt dieses Beispiel heute geradezu harmlos.[4]

Value-sensitive design meint hier nun zweierlei: Erstens eine grundsätzlich proaktive Haltung der Designer von Informationssystemen gegenüber der Erkenntnis, dass ihre Erzeugnisse wichtige Werte wie Autonomie, Eigentum, Fairness, Freiheit, Identität, informierte Zustimmung, Privatheit, Vertrauen, Wohlfahrt oder Würde beeinflussen. Zweitens eine Systematik, diese Einflüsse zu gestalten. Nötig ist hierbei begriffliche Arbeit (Welche Werte sind von einer bestimmten Technologie betroffen? Wie differenziert sich dieser Wert, z.B. Vertrauen, aus? Wer ist vom jeweiligen Wert direkt oder indirekt betroffen?), empirische Untersuchungen (Merken Nutzer, dass die Nutzung bestimmter Technologien einen *trade-off* von Werten – z.B. bequeme Nutzung vs. Privatheit – beinhaltet? Welche Unterschiede gibt es zwischen den Ansichten über bestimmte Handlungen und den tatsächlich durchgeführten Handlungen bei der Nutzung von Technologie?) und technische Analysen (Durch welche technischen Features drückt sich ein Wert in der Technologie aus? Wie kann ein gewünschter Wert – z.B. Kollaboration – durch das jeweilige Design der Technologie gefördert werden?). Wie eine solche Systematik konkret vonstatten geht, wird am besten anhand konkreter Fallstudien ersichtlich. Hier können solche Fallstudien aber aus Platzgründen nicht ausgeführt werden, die Leserschaft wird auf die Beispiele in Friedman et al. (2009) verwiesen.

Der mit Blick auf die Frage nach dem Guten in der Informatik interessante Punkt ist, dass das Gute nicht einfach «von aussen» an die Informatik herangetragen wird. Natürlich haben viele der von Friedman genannten Werte eine lange Geschichte und sind facettenreich ausdifferenziert. Doch diese Ausdifferenzierung ist eben auch eine Folge der konkreten Gestaltung und Nutzung unserer informationstechnischen Werkzeuge. Soziale Netzwerke

[3] Die nachfolgenden Ausführungen beruhen weitgehend auf Friedman et al. 2006 und Friedman & Kahn 2003.

[4] Informationssysteme wie Computer, Smartphones und dergleichen müssen in der Regel grosse Mengen an Daten generieren und speichern, um überhaupt funktionieren zu können, das Potenzial für Überwachung ist damit der Technologie quasi inhärent eingeschrieben. Das neue Buch des Sicherheitsexperten Bruce Schneier (2015) gibt hierzu eine konzise und leicht lesbare Einführung in das Ausmass der heutigen Massenüberwachung.

beispielsweise haben unseren Begriff dessen, was Privatheit bedeutet, verändert – und zwar nicht nur weil die Nutzer Sozialer Netzwerke generell schlicht unachtsam sind hinsichtlich des Teilens privater Information, sondern weil sie eben oft auch bewusst diese Informationen verbreiten, da sie denken, dass sich das gegenseitige Verständnis durch ein gewisses Mass an Aufbrechen von Privatheit verbessert. Somit funktioniert Technologie nicht nur einfach als Vermittlerin dieses Wertes, sondern sie prägt dessen Verständnis. Schliesslich sind Werte konkretisiert durch die Handlungen jener, welcher von den Werten betroffen sind. Und je mehr Informationstechnologie diese Handlungen mitgestaltet, desto enger verwebt sich «das Gute» mit der Theorie und Praxis der Informationsverarbeitung. *Value-sensitive design* strebt an, diese enge Verzahnung des Guten mit der «kalten» Technologie, materialisiert durch Hardware und Programmiercode, zu verdeutlichen – insbesondere auch bei jenen, welche diese Technologie bauen.

Value-sensitive design meint aber auch die bewusste Nutzung des Wissens, wie Technologiedesign und Werte miteinander verknüpft sind und sich gegenseitig bedingen. Hier geraten wir in ein heute kontrovers diskutiertes Themenfeld, was wir unter das Stichwort der «moralischen Technologien» fassen (*moral technologies*): Je mehr man darüber weiss, welche Faktoren das moralische Handeln und Verhalten von Menschen beeinflussen, desto stärker kann man dieses Wissen zur Einflussnahme auf das Verhalten nutzen. Verhaltensökonomen sprechen hier von *nudging*, man «stupst» die Nutzer auf den moralisch richtigen Weg.[5] Der Punkt ist nun, dass Informationstechnologie wie kaum eine andere Form der technologischen Intervention ein solches *nudging* ermöglicht: Indem technische Sensoren die Welt erfassen, verwandeln sich deren Phänomene in Bitströme, die ein viel breiteres Spektrum an Manipulation ermöglichen. Die Manipulation kann auch direkt in die Technologie eingebaut werden – unter Verwendung von Erkenntnissen des *value-sensitive design* über die Art und Weise, wie Technologie einen (positiv besetzen Wert) fördert.

Hierzu ein Beispiel: In Feldforschungen wurde ermittelt, welchen Effekt es hat, in einem Wohnquartier den durchschnittlichen Stromverbrauch pro Haushalt zu veröffentlichen. Es zeigte sich ein zweischneidiger Effekt: Mehrverbraucher werden in der Regel durch solche Veröffentlichungen motiviert, ihren Stromverbrauch zu senken – jene aber, die unter dem Durchschnitt liegen, verlieren oft die Motivation, ihre gute Stellung zu halten. Das kumulative Ergebnis: der Mittelwert ändert kaum. Nun könnte man im Wissen um diese psychologischen Mechanismen den Durchschnittswert manipulieren. Natürlich könnte man ganz krude einfach einen falschen Wert kommunizieren in der Hoffnung, dass auch «die Guten» weiterhin angehalten sind, ihre Anstrengungen zu erhöhen – würde allerdings eine solche Manipulation publik, wäre der Schaden bei künftigen derartigen Projekten immens. Die Manipulation könnte aber subtiler sein: Die Programme, welche den individuellen Stromverbrauch im *smart meter* des jeweiligen Haushalts messen, könnten das Sampling (d.h. über welchen Zeitraum der Durchschnitt ermittelt wird; das ist bis zu einem gewissen Grad immer eine willkürliche Entscheidung) so verändern, dass Sprünge weg vom Durchschnitt akzentuiert werden – etwa um das schlechte Gewissen bei starkem Mehrverbrauch zu verstärken. Die Systeme könnten «regulären» vom «irregulären» Minderverbrauch (etwa wenn die Leute einfach in den Ferien sind) unterscheiden und unterschiedlich kommunizieren. All dies kann auf der Ebene der Systeme selbst geschehen, auf eine Weise, die weder für die Nutzer noch für die Wartungsfachleute ersichtlich ist.

Informationssysteme könnten aber auch direkt zu Instrumenten werden, die den Nutzer selbst über seine eigenen Werte aufklären. Natürlich ist dieses Unterfangen nie alleine eine Sache der Informatik, sondern braucht massgeblich auch andere Disziplinen. Das online-Tool «Smartvote» beispielsweise nutzt Erkenntnisse der Politikwissenschaften, um politische Ansichten von Wählern mit jenen von Politikern zu vergleichen und damit als Wahlhilfe zu dienen. Es ist aber mehr als das: Das Instrument generiert eine «Bildsprache» in der Datenvisualisierung (die «smartvote-spiders»), die den politischen Diskurs mitgestaltet; politische Positionen werden in Diagramme übersetzt und entsprechend medial diskutiert, Rückkopplungseffekte entstehen (z.B. «lernen» Politiker, wie sie antworten müssen, um die gewünschte Visualisierung zu erreichen), und es entstehen vermutlich auch Prozesse der Selbsterkenntnis bei den Wählern (man fragt sich beispielsweise, ob die Visualisierung nun wirklich die eigenen Meinungen abbilden oder nicht). Wie die Diskussion rund um Smartvote zeigt (siehe dazu z.B. den Smartvote-Blog: *http://blog.smartvote.ch/?cat=10*), sind durchaus kritische Fragen angebracht – gleichzeitig zeigt das Beispiel aber

5 *Nudging* bezeichnet Interventionen, mittels derer Verhalten von Menschen auf vorhersagbare Weise beeinflusst werden kann, ohne dabei auf Verbote und Gebote zurückgreifen oder ökonomische Anreize verändern zu müssen. Grundlage dafür sind Erkenntnisse aus der Psychologie, Biologie, Verhaltensforschung, Sozialwissenschaft und anderen Bereichen über Mechanismen, die menschlichen Verhaltensweisen typischerweise unterliegen. Der Wirtschaftswissenschaftler Richard Thaler und der Rechtswissenschaftler Cass Sunstein haben mit ihrem 2008 erschienenen Buch «Nudge: Improving Decisions About Health, Wealth, and Happiness» die Debatte um das *nudging* angestossen, das auch viele kritische Repliken herausgefordert hat.

auch das Potenzial von Informationssystemen, die als Instrumente zur Selbsterkenntnis eingesetzt werden. Wir selbst arbeiten derzeit an der Idee, durch die Nutzung von Videospielen Instrumente zur moralischen Selbsterkenntnis zu generieren (mehr dazu in Christen et al. 2012) – dazu können wir zu einem späteren Zeitpunkt mehr berichten.

4. Informatik-Systeme als Schöpfer des Guten

Es ist eine Sache zu erkennen, dass Informationstechnologie das menschliche Verständnis des Guten – also Werte und deren Interpretationen – prägen und verändern und dass man diese Technologien im Sinne des *value-sensitive design* bewusst gestalten will. Eine doch recht verschiedene Sache ist es aber, wenn diese Systeme selbst das Gute in einem relevanten Sinne erkennen und danach handeln sollen – also zu *moral agents* werden könnten. Etwas derartiges zu schaffen, dürfte den Rahmen der Informatik wohl sprengen; es wird das Wissen und die Kompetenzen vieler Disziplinen benötigen – nicht nur technischer Disziplinen wie etwa die Robotik, sondern auch sozial- und geisteswissenschaftliche Disziplinen, zumal solche künstlichen *moral agents* in einer menschlichen Welt agieren würden (Sullins 2006).

Derartige Gedanken erscheinen zuerst einmal wie Science Fiction – und in der Tat gibt es zahlreiche Vorbilder in Literatur und Film, in denen Computersysteme eine Form von «Selbstbewusstsein» erhalten, das ihnen moralisches Denken und Handeln ermöglicht. Interessanterweise sind viele dieser Geschichten negative Utopien – negativ für den Menschen, der vom System gewissermassen als moralisch ungeeignet oder gar schlicht als Feind erkannt wird – wie der Computer HAL, der in «Space Odyssee 2001» den Menschen aus der Raumstation ausschliesst, um die Mission nicht zu gefährden, oder Skynet aus den Terminator-Filmen, wo der Computer zum Mittel des atomaren Erstschlags gegen die Menschen greift. In eine ähnliche Richtung zielen die Gedanken von Technik-Utopisten wie Ray Kurzweil, welche das Mooresche Gesetz der periodischen Verdoppelung der Rechenkraft von Computern zu dem Punkt extrapolieren, in dem diese «künstliche Intelligenz» uns sowieso überflügelt.

Doch lassen wir diese Utopien vorerst beiseite. Warum sollte die Informatik im Verbund mit anderen Disziplinen in Richtung von Systemen arbeiten, die das «Gute» quasi selbst erkennen können? Der Grund liegt in der zunehmenden Automatisierung, welche immer mehr Lebensbereiche erfasst: Je stärker diese Automatisierung in soziale Welten vordringt, desto mehr spielen bei den nötigen Entscheidungen auch moralische Gesichtspunkte eine Rolle. Bei den vollautomatisierten Fliessbändern der Automobilindustrie mögen solche Überlegungen noch irrelevant sein – wenn aber KI-Systeme wie «Watson»[6] dereinst anstelle des Versicherungsagenten ermitteln soll, ob eine gewisse Person eine Versicherung zu welchen Konditionen angeboten erhält, rücken moralische Fragen ins Zentrum.

Interessant ist hierbei insbesondere die Motivation der Entwickler solcher Systeme: Man will «Fehler» vermeiden – auch solche moralischer Art, also etwa ungerechtfertigte Diskriminierung und dergleichen. Analog wie in den Science Fiction-Szenarien wird der Mensch als moralische Fehlerquelle betrachten und das Informationssystem soll im besten Fall das moralische Dilemma gar nicht erst entstehen lassen. Bemerkenswert sind die Argumentationen pro autonome technische Systeme in Bereichen, wo tödliche Dilemmas auftreten können: autonome Fahrzeuge und autonome Kampfroboter. In beiden Fällen argumentiert man, dass das System seine Aufgabe zuverlässiger erledigt und damit das Auftreten moralisch problematischer Situationen – Unfälle oder Kollateralschäden, etwa als Folge einer psychischen Überbelastung von Soldaten – vermindert. Es manifestiert sich hier ein «Ingenieur-Blick» auf das moralische Problem an sich. An einer Konferenz erläuterte einmal der Technikethiker Jeroen van den Hoven, dass ein Ingenieur das berühmte Trolley-Problem[7] ganz anders sehen würde: Dieser würde fragen, wie man denn um Gottes Willen ein System habe bauen können, das zu einem derartigen Dilemma führe. Analog setzten die Entwickler des Google-Autos darauf, dass die Prognosefähigkeit eines autonomen Autos es weitgehend verhindern sollte, dass z.B. die Wahl zwischen dem Überfahren eines Menschen und dem Tod der Insassen überhaupt nötig wird.

Dennoch dürfte das moralische Dilemma die Messlatte sein, an der sich zeigen wird, auf welche Weise diese Systeme «das Gute» denn nun erkennen. Nicht wenige Forscher arbeiten derzeit an der Frage, wie man eine Art Moralsystem in Informationssysteme einprogrammieren kann (Wallach & Allen 2010). In-

6 «Watson» ist ein von IBM hergestelltes kognitives, lernendes Computersystem, das natürliche Sprache verarbeiten kann. Derzeit wird das System in zahlreichen Anwendungen geprüft, z.B. im Gesundheitswesen (Analyse publizierter Literatur zu Krankheiten zwecks Entwicklung neuer Therapien), in der Finanzwirtschaft oder in der Versicherungsbranche.

7 In diesem ursprünglich von der Philosophin Philippa Foot vorgeschlagenen Gedankenexperiment steht eine Person vor der Entscheidung, einen steuerlosen Zugwaggon, der auf fünf Gleisarbeiter zurast, auf ein Abstellgleis umzuleiten, wo allerdings eine andere Person arbeitet und ums Leben kommen würde. Varianten dieses Dilemmas sind das wohl bestuntersuchteste experimentelle Paradigma in der Moralpsychologie.

teressanterweise widerspiegeln die verschiedenen Ansätze die gängigen Ethik-Theorien: Eine «deontologische Strategie» zielt darauf ab, ein regelbasiertes System zu entwickeln, so dass das System je nach Situation auf eine geeignete Maxime zurückgreifen kann. Freilich ist hier die Schwierigkeit, dass kaum alle Situationen geeignet abgebildet werden können und dass ein Weg gefunden werden muss, wie das System eine abstrakte Regel auf ein konkretes Entscheidungsproblem anwenden kann. Eine «konsequentialistische Strategie» würde im System primär die Fähigkeit fördern, Handlungsalternativen simulieren zu können, so dass dann die Ergebnisse an einem bestimmten Zielwert bemessen werden können. Hier stellt sich unter anderem die Frage, inwieweit das System solche Alternativen überhaupt erst erkennen kann (auch Menschen haben hier ihre Mühe). Ein «tugendethischer Ansatz» würde primär auf die Lernleistung des Systems abzielen – das aktuelle Schlagwort hier ist «deep learning».[8] Welche dieser Strategien dereinst zu «moralischen Maschinen» führen wird, ist unklar. Nick Bostrom und Eliezer Yudkowsky (2014) erinnern aber daran, dass der menschliche Moralmassstab für das Design solcher Maschinen wohl nicht ausreichend sei. Die Geschichte zeige, dass viele einst als moralisch akzeptierte Verhaltensweisen heute als klar unethisch gelten würden – wer wisse schon, welche der heutigen Selbstverständlichkeiten in Zukunft gleichermassen auf moralische Ablehnung stossen würden. Wir sollten, so die Autoren, die Moral künftiger «Superintelligenzen» nicht durch unsere eigenen moralischen Vorurteile beschränken. Sollten derartige Informationssysteme einmal Wirklichkeit werden, dann hätte die Frage nach dem Guten in der Informatik tatsächlich eine unerwartete Wendung genommen: Die Disziplin würde eine Form des Guten schaffen, die wir uns derzeit noch gar nicht vorstellen können. ■

8 Gemeint sind damit neuronale Netzwerke von hoher Komplexität, denen man eine Unmenge von Daten füttern kann und die basierend darauf dann selbstständig Klassifikationsleistungen erbringen können, ohne dass man dem System sagt, worauf es achten soll. Solche Systeme sind auch erkenntnistheoretisch interessant, denn es entwickelt sich durch den Lernprozess ein derart komplexes System (im Wesentlichen die Verknüpfungen zwischen den künstlichen «Neuronen» und deren Gewichte), dass das System selbst seine Beschreibung darstellt; man kann es nicht durch einen Programmcode «zusammenfassen» und die Entwickler können es nach eigenen Aussagen selbst nicht mehr verstehen oder erklären.

Literatur

Bostrom N, Yudkowsky E (2014): The Ethics of Artificial Intelligence. In: Ramsey W, Frankish K (eds.) Cambridge Handbook of Artificial Intelligence. Cambridge: Cambridge University Press, 316–334.

Christen M, Faller F, Götz U, Müller C (2012): Serious Moral Games. Erfassung und Vermittlung moralischer Werte durch Videospiele. Zürich: Edition ZHdK.

Floridi L (2013): The Ethics of Information. Oxford: Oxford University Press.

Floridi L (1999): Information ethics: On the philosophical foundation of computer ethics. Ethics and Information Technology 1, 37–56.

Friedman B, Kahn PH Jr., Boring A (2006): Value Sensitive Design and Information Systems. In: Zhang P, Galletta D (eds.) Human-Computer Interaction in Management Information Systems: Foundations. New York: M.E. Sharpe, 348–372.

Friedman B, Kahn PH Jr. (2003): Human Values, Ethics, and Design. In Jacko JA, Sears A (Eds.) The human-computer interaction handbook. Mahwah, NJ: Lawrence Erlbaum Associates, 1177–1201.

Schneier B (2015): Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World. W. W. Norton & Company, New York.

Sullins JP (2006): When Is a Robot a Moral Agent? International Review of Information Ethics 6, 23–30.

Wallach W, Allen C (2010): Moral Machines: Teaching Robots Right from Wrong. Oxford: Oxford University Press.